15. (Amended) A method of determining whether a conferee in a videoconference is speaking, comprising analyzing whether visual lip movements of said conferee are reasonably consistent with an audio signal from a conference station in which said conferee is located such that the combination of lip movement and audio signal indicates human speech.

B3 D1

REMARKS

In the subject Office Action, all the pending claims 5 – 15 are rejected either under 35 USC§ 102(b) as being anticipated by Zhou (US Patent No. 5,512,939) or under 35 U.S.C. §103(a) over Ogata, et al. (JP 06-062400), Kamata, et al. (U.S. Patent No. 5,953,050) in view of Zhou (US Patent No. 5,512,939), as well as under the judicially created doctrine of double patenting U.S. Patent No. 5,914,747 of the same inventor in view of Zhou. The applicant has further amended independent claims 5, 11, 14 and 15 to more clearly and distinguishably define the invention and respectfully traverses the rejections based on the amended claims, as explained in detail below.

The present invention teaches a novel technique applicable in videoconference environment to precisely and affirmatively determining who is speaking. In particular, as now precisely and explicitly defined in the amended independent claims 5, 11, 14 and 15, the present invention teaches a determination of whether a conferee is speaking by analyzing whether the lip movement of the conferee is reasonably consistent with the audio signal from the conference station where the conferee is located so as to produce human speech. In other words, the conferee is determined to be speaking if the lips of the conferee are moving in a way that is reasonably

consistent with the audio signal transmitted by the conference station that the conferee is located (page 5, lines 3-4). Thus, in the present invention, merely finding of a coexistence of lip movement of the conferee and the audio signal from the conferee's station is not sufficient to determine that the conferee is speaking. It is the correlation, not merely the coexistence, between the lip movement and the audio signal associated with the conferee determines that the conferee is speaking. If a conferee were chewing while music is being played, the system of the present invention would NOT determine that this conferee was speaking. However, if both the meter and the form of lip movement corresponded to the pattern of audio, it is determined that the conferee is speaking. This distinguishing feature realizes a precise and affirmative determination on whether the conferee is speaking.

The applicant respectfully submits that this distinguishing feature is neither disclosed nor implied in Zhou (US Patent No. 5,512,939) or other cited patents.

In particular, as taught throughout the disclosure in Zhou, whether the conferee is speaking is determined merely by a coexistence of lip movement and an audio signal at the station where the conferee is located. In other words, as long as there is a movement of the lips of the conferee at the same time as there is an audio signal from the conferee's station, the conferee is determined likely to be speaking. This can not be precise and affirmative because 1) the movement of the lips does not always generate speech, and 2) the audio signal in the conferee station does not always result from the conferee's speech. For example, the conferee may be falsely determined as being speaking if the conferee is yawning or eating something while there is music being played in the conferee's station. Therefore, a determination made merely by the coexistence between the lip movement and the audio signal is not precise and affirmative. In fact, as described throughout the disclosure of Zhou, when the lips of a conferee are moving at the same time when the audio signal is detected, the conferee can only be determined to be "most

likely" talking (see col. 1, line 66 – col. 2, line 1; col. 14, lines 45-48; col. 15, lines 39-44; col. 16, lines 56-65). In Zhou, for purpose of determining whether the conferee is speaking, the lips of the conferee can be moving in any way as long as they are found to be moving at the same time there is an audio signal detected. No correlation between the lip movement and the audio signal has been suggested or implied in Zhou to determine that the conferee is speaking.

Both Ogata and Kamata use only audio signal from a conference station to determine whether the conferee in the station is speaking. Thus, a combination of Ogata and/or Kamata with Zhou can not conclude the present invention. Thus, the applicant believes that the amended independent claims 5, 11, 14 and 15 are not obvious over Zhou, Ogata, Kamata and/or their combination, and are therefore patentable. At least for the same reasons, their independent claims 6 – 10 and 12- 13 are also patentable.

Furthermore, the applicant does not agree that the present invention as claimed is obvious over the patent 5,914,747 of the present inventor in view of Zhou. In particular, the distinguishing feature that whether the conferee is speaking is determined by analyzing the correlation between the conferee's lip movement and the audio signal from the conferee's station is neither disclosed in the claims of the patent 5,914,747 or in Zhou as explained above. The 5,914,747 patent is directed to conservation of bandwidth in the case where no conferee is present in a particular conference station. Zhou, as discussed above, establishes a likelihood of speech, not a clear determination of human speech by analysis of a correlation between lip movement and the audio signal. The obviousness double-patenting rejection is thus respectfully traversed.

Applicant believes that the application as amended above is now in good condition for allowance, and reconsideration is here respectfully requested in view of the amendments and the above remarks. A replacement of the amended Figure 1 is also enclosed here as required. The

Examiner is authorized to deduct additional fees believed due from our Deposit Account No. 11-0223.

Respectfully submitted,

KAPLAN & GILMAN, L.L.P.
900 Route 9 North
Woodbridge, New Jersey 07095
Telephone (732) 634-7634

DATED: May 8, 2002

Jeffrey I. Kaplan
(Reg. No. 34,356)

CERTIFICATE OF MAILING

I hereby certify that this correspondence is being deposited with the United States Postal service as first class mail, in a postage prepaid envelope, addressed to Box RCE, Commissioner for Patents, Washington, D.C. 20231 on May 8, 2002.

Dated May 8, 2002

Signed _____

Print Name Jeffrey Kaplan

Marked-up version of amended claims 5, 11, 14 and 15

(Amended) A videoconferencing system comprising:

a conference bridge for interconnecting a plurality of remotely located videoconference stations;

means for determining whether a conferee is speaking by analyzing [a consistency between] **whether** [a] visual lip movement**s** of said conferee [and] **are reasonably consistent with** an audio signal from a conference station in which said conferee is located **so as to produce human speech**; and

means for visually altering an image of said conferee displayed in other conference stations if said conferee is determined to be speaking.

11. (Amended) A videoconference station comprising:

a transmitter to transmit a combined audio video signal to a videoconference bridge; and

means for determining whether a conferee located at said videoconference station is speaking by analyzing [a consistency between] **whether** [a] visual lip movement**s** of said conferee [and] **are substantially consistent with** an audio signal at said station **so as to indicate human speech**.

14. (Amended) A method of displaying images of a plurality of conferees in a videoconference system, comprising:

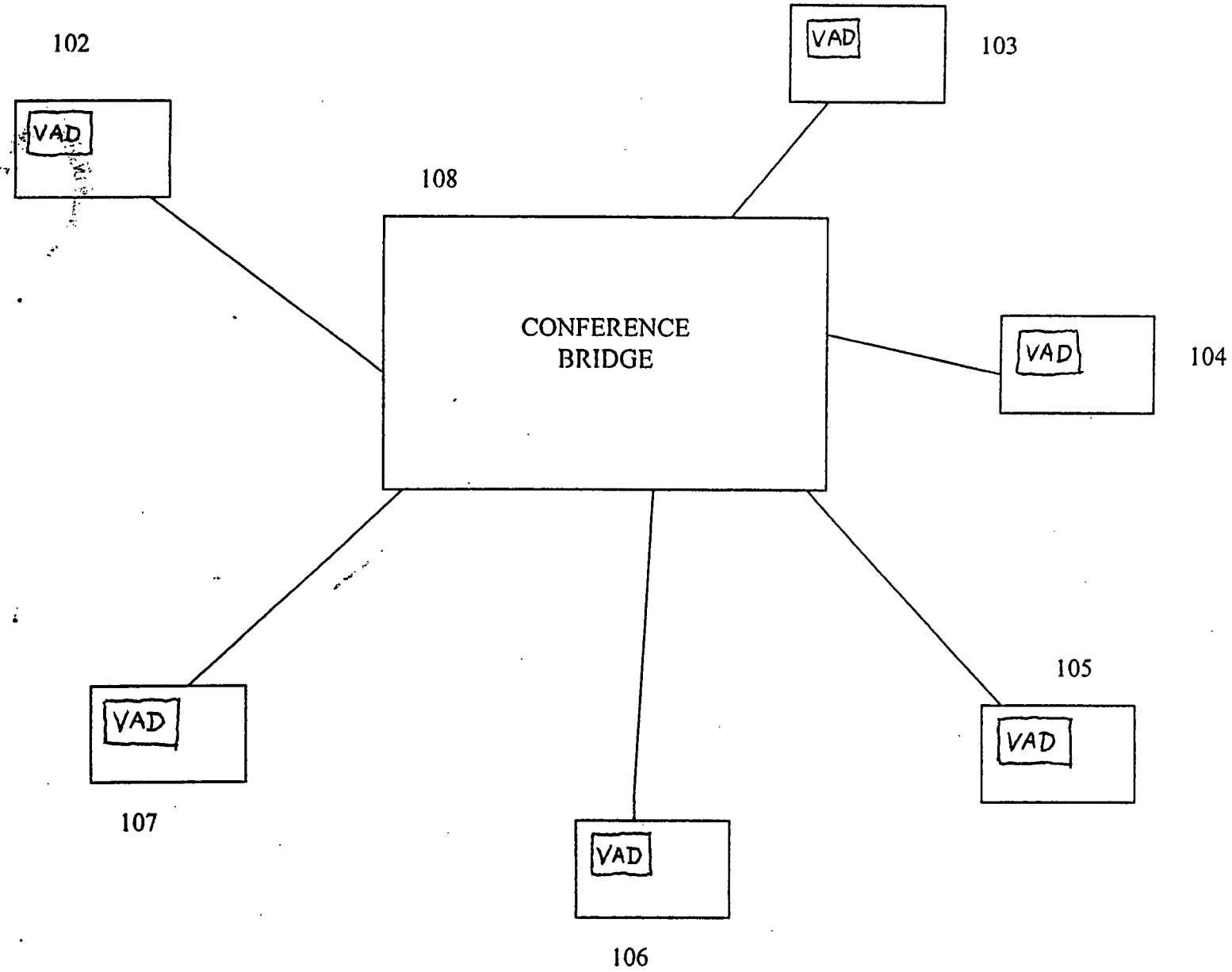
determining whether a conferee is speaking by analyzing a consistency between [a] visual lip movement**s** of said conferee and an audio signal from a conference station

in which said conferee is located **such that the combination of lip movement and audio signal indicates human speech**; and

visually altering an image of said conferee that is displayed to other conferees when said conferee is determined to be speaking.

15. (Amended) A method of determining whether a conferee in a videoconference is speaking, comprising analyzing [a consistency between] **whether** [a] visual lip movements of said conferee [and] **are reasonably consistent with** an audio signal from a conference station in which said conferee is located **such that the combination of lip movement and audio signal indicates human speech**.

FIG. 1


102
VAD
103
VAD
108
CONFERENCE
BRIDGE
VAD
104
105
VAD
VAD
107
VAD
106